SELECTION OF SPATIAL AND TEMPORAL DISCRETIZATION FOR GROUNDWATER AND OVERLAND FLOW MODELS

A. M. Wasantha Lal [1]

Spatial and temporal discretizations are key factors deciding the optimal use of computer resources in groundwater and overland flow models. The discretization should be sufficiently fine to describe the solution with a reasonable resolution, and prevent excessive numerical errors. An excessively fine spatial and temporal discretization could be very expensive because of the computer storage cost and the running cost. The optimal resolution of a model also depends on the intended use of the solution, and the types of stresses on the model. The study is aimed at understanding numerical errors and run times of overland and groundwater flow models simulating the affects of internal and boundary stresses caused by well pumping, canal level fluctuations, rainfall, etc. under steady and unsteady conditions. The results are used to understand and apply existing and new groundwater and overland flow models to simulate hydrology at regional and local scales in South Florida.

Fourier analysis of the linearized governing equation is used in the study to obtain analytical expressions for numerical errors and run times. Numerical experiments carried out with the MODFLOW model (McDonald and Harbough, 1984), and a number of implicit and explicit models show that the analytical expressions can be used to compute errors and run times accurately. These expressions use dimensionless parameters so that the results can be used for other modeling applications too. Examples are shown to demonstrate the use of the results.

[1]Lead Engineer, South Florida Water Management District, 3301 Gun Club Rd., West Palm Beach, FL 33406

INTRODUCTION

The number of computer models used to simulate various overland flow and ground water flow conditions has increased recently due to the increased environmental, agricultural and developmental interests. South Florida is one such area in which models of different scales are used for planning, management and regulation of water resources. In South Florida, regional models are used mostly in planning and management of water resources while medium and small scale models with county-wide and local coverages are used in regulatory and permitting functions. The multi-agency efforts to implement the restoration of the Everglades also has increased various modeling efforts. As a result of the multiple and overlapping use of models, interest has recently grown over understanding and proper application of them, and the interpretation of their results. The current study is aimed at understanding the relationship of spatial and temporal discretizations to the numerical errors and run times. Both steady and unsteady cases are investigated. South Florida is used as a test area for many of the methods developed.

Most of the ground water and overland flow models are developed around a numerical method solving the parabolic partial differential equation which is sometimes referred to as the diffusion equation. Diffusion flow models of varying resolutions are used to look into hydrologic processes at different scales. Numerical models of any scale contain uncertainties due to uncertainties in the input, the parameters, and the algorithms. Input uncertainty is due to inaccurate or sparse spatial and temporal input data such as rainfall, and evapotranspiration. It can be reduced by improving the density of the data collection network and increasing the data quality. Parameter uncertainty is mainly due to inaccurate parameter values used to represent spatially varying physical characteristics. This error can be reduced somewhat by calibration (Lal, 1995). Numerical errors are considered to be the source of algorithm uncertainty in the study. Various unconditionally stable numerical methods using implicit or other methods have made it possible for modelers to use almost any discretization with computer models. Unlike in explicit methods where there is some error control because of the stability condition, implicit models such as MODFLOW need guidelines to select discretizations so that the error becomes known and controllable.

Many rules of thumb have been used in the past to determine both spatial and temporal discretizations so that the physical domain and the time variations of the solution can be represented sufficiently accurately. There have also been model applications in the past in which the discretizations are decided simply by the availability of data without regard to numerical considerations. Initially, model discretization was studied simply by analyzing the truncation errors. Based on these studies, Richtmyer and Morton (1967) compiled many of the basic developments behind consistence, convergence and stability of parabolic and other problems. In many of the early applications, the primary method of numerical error control is by adhering to a discretization satisfying the stability conditions derived using Von Neuman and other methods. Run time and numerical error were not directly used to formulate discretizations in many early applications.

Error analysis of partial differential equations was initially limited to an order-of-magnitude analysis. Error control is however commonly used in solving initial value problems involving ordinary differential equations, as in the Runge-Kutta-Fehlberg method and the Adams variable step-size predictor-corrector method (Burden and Fairs, 1985). To control numerical errors in MODFLOW applications, Anderson and Woessner (1991) suggested empirical methods based on model convergence. Hirsch (1989) used a method for error analysis based on linearization and Fourier Analysis. This method which is similar to Von Neuman method for stability analysis has been used for diffusion and other equations. Lal (1998) used the same method with additional expressions derived for computational time to evaluate and compare the computational performances of various numerical models used to solve diffusion equations. The subject of error analysis and output evaluation has become increasingly important because the space and time discretizations used in many model applications are often not small as required by the finite difference formulations. The use of unconditionally stable implicit methods has also complicated the use of stability condition as an error control. The current study extends the ideas of Fourier analysis used by Hirsch (1989) and Lal (1998) to develop expressions for numerical errors of many ground water flow and overland flow models.

Numerical errors are introduced when the solution to the governing partial differential equations is represented by discrete values in the model, and when these discrete values are used in numerical computation related to the finite difference method. Numerical errors are classified into three different types in this paper. Stresses and errors due to conditions such as changing water levels in canals, changing pumping rates in wells, and changing rains are analyzed separately in the study. Principle of superposition makes it possible to combine these results and apply to many practical problems. The error analysis in this study is carried out for an arbitrary Fourier component and the steady state. The results are presented in dimensionless forms so that they can be used in a variety of problems. They are verified using MODFLOW and other models. They include expressions for computer run time, numerical error and data storage requirements in terms of various discretization parameters. Results of the study are useful in understanding the output of existing models, and providing guidelines for designing future models.

NUMERICAL SOLUTION OF THE DIFFUSION EQUATION

Two dimensional groundwater flow and overland flow can be explained using the following governing equation. For overland flow, the equation is derived by neglecting the inertia terms in the St Venant equations. (Hromadka and Lai, 1985, Lal, 1998).

$$\frac{\partial H}{\partial t} = \frac{\partial}{\partial x}\left(K\frac{\partial H}{\partial x}\right) + \frac{\partial}{\partial y}\left(K\frac{\partial H}{\partial y}\right) + \frac{S}{s_c} \tag{1}$$

in which, $H = h + z$ = water level or water head; S = source and sink terms representing rainfall, evapotranspiration and infiltration. For overland flow, h = water depth; $K = \frac{h^{\frac{5}{3}}}{n_b\sqrt{S_n}}$ when the Manning's equation is used; n_b = Manning's coefficient; S_n = water surface slope and $s_c = 1$. For groundwater flow, s_c = storage coefficient; $K = T_c/s_c$ in which T_c = transmissivity of the aquifer, assuming an isotropic material; $T_c \approx k_c\bar{h}$ for unconfined flow in which $\bar{h}$ = water depth of the saturated layer and k_c = hydraulic conductivity. In the case of groundwater flow, K is used in place of T_c/s_c to simplify the derivations. The flow vector is computed using

$$\vec{Q} = Ks_c\vec{\nabla}H \tag{2}$$

in which, $\vec{Q}$ = flow rate per unit width. When a weighted implicit finite volume formulation is used, (1) can be expressed for an arbitrary cell as (Lal, 1998)

$$H_{i,j}^{n+1} = H_{i,j}^{n} + \alpha\, Q_{net}(H^{n+1})\frac{\Delta t}{s_c \Delta A} + (1-\alpha)\, Q_{net}(H^{n})\frac{\Delta t}{s_c \Delta A} + \frac{\bar{S}\Delta t}{s_c \Delta A} \tag{3}$$

in which ΔA = area of the cell; Q_{net} = net inflow to the cell; α = weighting factor for semi-implicit schemes; n = time step; $\bar{S}$ = weighted average source term for the area during the period. For a rectangular cell, Q_{net} is given by

$$\begin{aligned} Q_{net} = {} & K_{i+\frac{1}{2},j}(H_{i+1,j} - H_{i,j}) + K_{i-\frac{1}{2},j}(H_{i-1,j} - H_{i,j}) \\ & + K_{i,j+\frac{1}{2}}(H_{i,j+1} - H_{i,j}) + K_{i,j-\frac{1}{2}}(H_{i,j-1} - H_{i,j}) \end{aligned} \tag{4}$$

Explicit and the implicit methods are obtained by using $\alpha = 0$ and 1.0 with (3) and (4).

NUMERICAL ERROR ANALYSIS

Numerical errors are present in computer models because of their inability to represent or solve continuous functions without digitizing or discretizing them. In the paper, numerical errors are classified into three categories.

Types of numerical errors

Numerical errors are classified into the following three categorized based on the ways in which they are introduced into a solution.

A A numerical error is introduced when the initial and boundary condition data are recorded and provided to the model as discrete values in time and space. Due to this type of error in "representation", frequencies in the solution above a certain value are either completely left out, or not represented accurately.

B The second type of error is introduced when the internal discretization within a model is not sufficient to carry the solution accurately over the entire time and the space. To understand this error, consider a case in which the water level boundary condition data are adequately represented in time. If the spatial discretization is not sufficient to represent the resulting disturbance in space along its entire travel path, an error of this type is introduced to the

solution. Similarly, even if the initial condition is represented accurately in space, a coarser time step may introduce errors after some time.

C The computational error introduced during numerical computations carried out using the left over frequencies in the model after steps (A) and (B) is referred to as the third type of error. This type of error is analyzed using methods spectral methods used by Von Neuman.

Higher frequency components are subjected more to all three types of errors. Error types (A) and (B) are introduced in general due to the inability of the model to represent continuous solutions accurately using discrete values. These errors can be quantified using an arbitrary Fourier component of the solution. Assuming that a component can be described using its wave number k ($k = 2\pi/$ wavelength) or frequency f ($f = 2\pi/$ period), the following approximate expressions were obtained for maximum percentage discretization errors of 1-D and 2-D problems. A Monte Carlo method fitting different 1-D and 2-D wave shapes to a grid was used to obtain them.

$$\phi = 0.5\sqrt{\epsilon_d} \quad \text{or} \quad \epsilon_d = 4.0\,\phi^2 \quad \text{for 1-D} \tag{5}$$

$$\phi = 0.35\sqrt{\epsilon_d} \quad \text{or} \quad \epsilon_d = 7.8\,\phi^2 \quad \text{for 2-D} \tag{6}$$

in which, ϵ_d = maximum percentage error in discretization measured at the center of the cell; $\phi = k\Delta x$ is the dimensionless forms of Δx. The same equation applies when ϕ is replaced with ψ in which $\psi = f\Delta t$. Quantity ϕ was also used by Hirsch (1989) to make Δx dimensionless. ϵ_d only depends on the geometrical shape of the wave form. For 1-D problems, 1% and 5% errors in discretization for example correspond to ϕ or ψ = 0.5 and 1.1 respectively. For 2-D problems, they correspond to ϕ or ψ = 0.35 and 0.80 respectively. An easier way to visualize ϕ or ψ is to consider that approximately $\frac{\pi}{\phi}$ grid spaces or discretizations are needed to describe half the wave length of a sine wave. It can be seen that approximately six grid spaces are needed over the length of half a sine wave to represent it with a maximum error of 1%. Three discretizations per half sine wave or ϕ = 1.05 brings the error upto 4.5%. Equation (6) can also be obtained using actual model runs (Lal, 1998). Error ϵ_d is the smallest maximum error possible with a model using a discretization ϕ and any time step.

The error explained in (B) can be understood by realizing that k and f of a single Fourier component are related as a result of the governing equations. If a Δx is selected so that a certain wave number k describing the initial condition is represented with a known accuracy, Δt also has to be selected so that f is represented with a comparable accuracy. Similarly, if Δt is selected to represent a certain frequency f of the boundary condition data, k in the solution also has to be represented with a reasonable accuracy. Once a discretization is selected to represent a certain Fourier component with a given accuracy, lower frequency components are automatically represented with a higher accuracy, and do not need special attention. Discretizations are generally designed for the highest frequency or the wave number of interest. If the ϕ, and therefore the accuracy is known for a certain component with a wave number k, then ϕ' value for a different wave number k' can be determined using $\phi' = \phi k'/k$. If, for example, a discretization describes a 100 m wave with an accuracy of 1%, the same discretization can describe a 45 m (=100*0.5/1.1) waves with an accuracy of 5%. The accuracy of a numerical computation cannot be better than the accuracy at which f or k in the solution is represented. If a higher accuracy is needed from a computational scheme, the input data should be prepared with smaller ϕ and ψ values corresponding to the expected high accuracy.

Closely matching pairs of Δx and Δt can minimize errors of type B. In order to compute an error of type (B), a relationship between f and k can be obtained first for diffusion flow using solutions of the form $H = H_0 e^{I(kx-ft)}$ and $H = H_0 e^{I(kx+ky-ft)}$ respectively for 1-D and 2-D problems in which $I = \sqrt{-1}$. In the case of 2-D problems, k is assumed to be the same in both x and y directions for simplicity. The relationship obtained can be expressed as

$$f = dKk^2 \tag{7}$$

in which, $d = 1$ and 2 for 1-D and 2-D problems respectively. Numerical errors of type (B) are created if both f and k of the solution are not represented accurately in the model.

Computational errors

Computational errors explained in (C) are estimated using a methods similar to that used in the Von Neuman stability analysis. In the analysis, the behavior of the numerical scheme in response to an arbitrary i th harmonic with wave number $k_i = \frac{i\pi}{N}$ is compared with the behavior of the governing equation with respect to the same harmonic. The wave number is defined as $2\pi/\lambda$ in which λ = wave length. Assuming a solution domain length of length L in x direction, a mesh spacing of Δx allows a minimum wave length λ_{min} of $2\Delta x$ and a maximum wave length of $2L$. In the analysis, a term ϕ_i defined as $\phi_i = k_i \Delta x$ is used to represent the i th harmonic in dimensionless form (Hirsch, 1989). The subscript is often removed for simplicity. A term $\psi = f\Delta t$ can be defined similarly to represent a harmonic in the time domain, in which the frequency $f = 2\pi/T_p$, and T_p = wave period. ϕ is used as the dimensionless variable to describe the spatial discretization.

For numerical methods based on finite difference methods, an analytical expression can be derived for the numerical error (Hirsch, 1989, Lal, 1998).

$$\epsilon = 1 - |G| \tag{8}$$

in which, ϵ = numerical error per time step as a fraction of the amplitude; G = ratio of amplitudes of numerical and analytical solutions, or the amplification factor of the numerical method.

$$G = \frac{1 - 4d(1-\alpha)\beta \sin^2(\phi/2)}{1 + 4d\alpha\beta \sin^2(\phi/2)} \frac{1}{e^{-d\beta\phi^2}} \tag{9}$$

$\beta = \frac{K\Delta t}{\Delta x^2}$ = non-dimensional form of Δt; $d = 1, 2$ for one and two dimensional problems with square grids. Equation (8) can be expanded to give

$$\epsilon = \pm\frac{d^2\beta^2\phi^4}{2} + \frac{d\beta\phi^4}{12} + \ldots = \pm\frac{d^2k^4K^2\Delta t^2}{2} + \frac{dKk^4\Delta t\Delta x^2}{12} + \ldots \tag{10}$$

in which $+$ and $-$ signs correspond to implicit and explicit models respectively. The cumulative numerical error after many time steps, ϵ_T, depends on the number of time steps n_t, and the error at each time step ϵ. Error ϵ_T is bound by $n_t\epsilon$, in which $n_t = T/\Delta t$.

$$\epsilon_T \approx \frac{\epsilon}{\beta\phi^2} TKk^2 = \frac{\epsilon}{d\beta\phi^2} fT \tag{11}$$

where, k = wave number of the harmonic; T = maximum duration over which a given harmonic stays in the computational domain and accumulates errors. Examples shown below demonstrate how fT is computed. In the case of rainfall driven stationary water level variations, $fT = \pi/4$ because the error is largest after a quarter cycle. In the case of a traveling disturbance generated by a change in water level, the disturbance travels at a speed of f/k, and covers a distance X in time T making $fT = kX$. It is shown later that the absolute error in the problem is maximum when $fT = 1$. Similarly it can be shown that $fT < 3$ for most practical applications.

Equation (11) can be simplified if ϕ is smaller than 1 by using a truncated Taylor series expansion. For explicit, implicit and semi-explicit 1-D and 2-D finite difference models,

$$\epsilon_T \quad \text{(expl/impl 1-D)} \quad \approx \quad \frac{fT\phi^2}{2}(\mp\beta - \frac{1}{6}) \tag{12}$$

$$\epsilon_T \quad \text{(semi-impl 1-D)} \quad \approx \quad fT\left[\frac{\phi^2}{12} - \frac{\phi^4}{12}(\beta^2 - \frac{1}{30})\right] \tag{13}$$

$$\epsilon_T \quad \text{(expl/impl 2-D)} \quad \approx \quad fT\phi^2(\pm\beta - \frac{1}{12}) \tag{14}$$

$$\epsilon_T \quad \text{(semi-impl 2-D)} \quad \approx \quad fT\left[-\frac{\phi^2}{6} + \frac{2\phi^4}{3}(\beta^2 + \frac{1}{120})\right] \tag{15}$$

The positive and negative signs apply for the explicit and implicit methods as shown. Semi-implicit methods use $\alpha = 0.5$. Explicit 1-D and 2-D models additionally require $\beta < 0.5$ and $\beta < 0.25$ respectively. Numerical experiments will later show that offsets of β such as $1/6$ and $1/12$ in (12) and (13) can be neglected specially with implicit methods using relatively large β. Above equations also show that semi-implicit methods are second order accurate in time.

Useful ranges of discretization parameters

Equation (11) is useful in computing numerical errors for any discretization, even if Δx and Δt are selected independently. However, matching pairs of Δx and Δt, and their upper bounds become useful when selecting the final values for practical applications. A useful but subjective upper bound for ϕ or ψ in 1-D problems is 1.6 for which the error in representing a sinusoidal component is approximately 10%. The upper bound for ψ places an upper bound for β which

can be obtained using the following equation derived using (7).

$$\beta = \frac{\psi}{d\phi^2} \tag{16}$$

The upper bound of β is obtained for a given ϕ by using $\psi = 1.6$ in (16). Figure 1 shows how the upper bounds $\phi = 1.6$ and $\psi = 1.6$ place limits on the applicable range of the error equation (11). If both k and f are to be represented with equal accuracy in a model, the corresponding β for the "matching condition" is given by (16) with $\phi = \psi$. The relationship between Δx and Δt under this condition is given by

$$\frac{\Delta x^2}{\Delta t} = dK\frac{\phi^2}{\psi} \tag{17}$$

It is unlikely that many practical discretizations would support solutions in time and space equally accurately. If the time step used in a model is larger than the matching value given by (17), the model will represent f with a smaller accuracy corresponding to $\psi = f\Delta t$. If the time step used is extremely large such that ψ exceeds the upper bound of approximately 1.6, the specific Fourier component will not be represented accurately at all, and only frequencies lower than f corresponding to an acceptable value of ψ will be represented in the model. This may sometimes create too much noise and trigger larger oscillations, specially with α close to 0.5. Under this conditions, either Δt has to be reduced to accommodate the frequency component, or Δx has to be increased to avoid it and save data space. If the time step used in a model is smaller than that given by (17), frequencies above $f = d\phi^2 K/\Delta x^2$ can be represented using the time step. But since there are no such frequencies, part of the data storage space is wasted. This space can be saved by reducing the spatial discretization to $\Delta x = \sqrt{(d\Delta t\ K\phi^2/\psi)}$ derived using (17). In these examples, discretizations were matched to make sure that resources are not wasted by selecting unmatched ϕ and ψ.

In the previous discussions, it was demonstrated that at least 6 grid points are needed to represent a half-sine wave in the model with 1% accuracy. However, models may often have rainfalls or stresses acting on one to very few cells at a time. When this happens, the model acts more like a lumped parameter model for the area, and the solution become grid dependent. Under

the condition, the solution may not accurately represent the flow as described by the governing partial differential equations, and the analytical expressions for numerical error become somewhat inaccurate. Consequently, the model will filter out some of the high frequency components it cannot represent, but still show many of the important physical behaviors. Such a model has limited uses, but only after a proper verification. Model results under such conditions have to be interpreted very carefully.

An exception applies if the cells in the model represent actual physical cells in the field and the equations governing the flow across cells in the model also represent actual physical conditions. With correct rainfall and ET in the cells, the model now represent the physical system accurately, even if it does not solve the diffusion equation. This situation is true in parts of South Florida where large ponded parts of wetlands or agricultural areas are surrounded by levees.

LATERAL PROPAGATION OF ERRORS

When the water level in a canal, tidal bay or the ocean changes, a disturbance in head is created which travels outward from the source of the disturbance. Water level changes due to such stresses constitute an important part of the solution in many models. Errors in the solution of these disturbances are also important. To understand numerical errors in such solutions, a sinusoidal disturbance $H = H_0 \sin(ft)$ is studied in 1-D. It can be shown that the analytical solution for head in such a propagation is

$$H(x,t) = H_0 e^{-kx} \sin(ft - kx) \tag{18}$$

in which, $f = 2Kk^2$. The analytical solution for discharge is given by

$$Q(x,t) = \sqrt{2} K k H_0 e^{-kx} \sin\left(kx - ft - \frac{\pi}{4}\right) \tag{19}$$

These equations show that the amplitude becomes less than 1%, 5% and 37% of the starting amplitude when $fT = kX >$ 5, 3 and 1 respectively. This shows that the waveforms become negligible after traveling close to one cycle. It also leads to the following expression which is similar to an expression derived by Townley (1995) and used by Haitjema (1995) for transient

state analysis.

$$\frac{X^2}{KT_p} = \frac{(\ln \alpha_d)^2}{2\pi} \quad \text{or} \tag{20}$$

$$\frac{KT}{L_p^2} = -\frac{\ln \alpha_d}{8\pi^2} \tag{21}$$

in which, T_p = period of the wave; L_p = wave length; α_d = decayed amplitude as a fraction of the original. These equations can be used to determine the length and time scales of one-dimensional diffusion wave disturbances of period T_p or wave length L_p respectively. They can be used as indicators to determine when the amplitude is so small that it makes no sense to compute the error as a percentage of the amplitude for the problem. The numerical error at a distance x from the line of disturbance is computed using (11) and $fT = kX$ as discussed earlier.

$$\epsilon_T(x) = \frac{k\epsilon}{\beta\phi^2}x \tag{22}$$

This equation shows that as a percentage, the numerical error increases linearly with x. Since water level fluctuations decrease exponentially, the absolute error increases and decreases with the distance, giving a maximum ϵ_T of approximately $0.37\phi^2\beta$, at $fT = kX = 1$.

The gradient of the ϵ_T versus x line given by $k\epsilon/(\beta\phi^2)$ can be used to obtain ϵ for various computer models. These ϵ values can be compared with the same values obtained using analytical expressions developed in the study. When estimating ϵ_T for numerical models, (18) and (19) are used as exact solutions. These exact values are subtracted from model values to compute numerical errors in heads and discharges. The peak and trough values of the sinusoidal variations have the largest errors. The exact values of ϵ for explicit, semi-implicit and the fully implicit MODFLOW models are obtained using (8) with α = 0.0, 0.5 and 1.0 with $d = 1$. In preparing the experimental setup to compute ϵ_T, the number of grid cells in the direction of propagation was limited to 100 to prevent excessive run times. Over 200 cycles of sine waves were generated for each experiment. Since dimensionless parameters are used, the actual physical dimensions and physical constants used in the tests are not important.

Figures 1, 2 and 3 show the ϵ values computed for the ADI, explicit and the MODFLOW(PCG2) models respectively. A range of ϕ values such as 0.2, 0.4 and 0.8 were used in the experiments. All the figures show that the analytical and numerical plots of ϵ agree very closely, implying that the analytical expressions derived for numerical error are accurate. These results are similar to the results shown by Lal (1998a) using a water level subsidence experiment. Figure 2 shows in dashed lines that the error measured as the (numerical value – analytical value) is small when $\beta \approx 0.16$, and becomes negative when $\beta < 0.16$. Figure 2 also shows that the approximate form of the analytical solution in (12) based on a truncated Taylor series is also relatively accurate. Figure 4 shows the variation of the amplitude and the numerical error with distance for the MODFLOW model. This figure demonstrates how the percentage error increase and the amplitude decrease with distance. The behavior of error with β and ϕ for a problem with a triangular mesh is similar as demonstrated in the paper by Lal (1998b).

NUMERICAL ERRORS OF FLOW VELOCITY AND DISCHARGE

In overland flow and groundwater flow models based on the diffusion equations, discharge across two neighboring cells is

$$Q^n_{i+1/2,j} = Ks_c \frac{H^n_{i+1,j} - H^n_{i,j}}{\Delta x} \tag{23}$$

in which, $H^n_{i,j}$ and $H^n_{i+1,j}$ are the heads of the cells; $Q^n_{i+\frac{1}{2},j}$ = flow rate betweens cells per unit width. In order to compute the numerical error in the flow, a solution of the form $H^n_i = E^n \exp(I\phi i)$ is substituted in (23) to obtain $Q_{i+1/2} = 2Ks_c E_n I \sin(\phi/2)/\Delta x$. Using an analytical solution of the form $H(x,t) = H_0 \exp(-Kk^2 t)\exp(Ikx)$ in the governing equation, it can be shown that $Q(x,t) = Ks_c kIH(x,t)$. The ratio between $Q^n_{i+1/2}$ and $Q(x,t)$ can now be used to compute the approximate numerical error as

$$\epsilon_Q = (1-G)\frac{2\sin(\frac{\phi}{2})}{\phi} \tag{24}$$

in which ϵ_Q = numerical error in discharge for one time step, as a fraction of the total discharge for the specific Fourier component. Numerical error in flow velocity and discharge are given by the same expression. Equation 24 shows that the error in head and discharge are approximately the same, with the former slightly higher.

The accuracy of (24) can be verified in the same way it was done for the head, by simulating the propagation of sinusoidal disturbances in head, and observing the decay of the peak values of the sinusoidal discharge rate with distance. Numerical error in the peak rate is computed by first running a model for a long period of time, passing over 10000 cycles of waves until traces of the initial condition no longer remain in the solution. Errors are computed by assuming the analytical solution (19) to be exact. The gradient of the error versus distance curve is used as before to compute ϵ_Q for the model. Figure 5 shows a plot of ϵ_Q with β for the model when $\phi =$ 0.5, 1.0, and 1.5. According to the figure, the analytical estimates of error compare well with the numerical estimates.

NUMERICAL ERRORS NEAR WELLS UNDER VARIABLE PUMPING RATES

Numerical errors of model results are large close to groundwater wells because of the extreme curvature in the solution. Thiem equation provides an approximate but efficient method to compute water levels very close to a well when the water level of the cell is known (Anderson and Woessner, 1991). Numerical error close to a well subjected to a variable pumping rate is investigated in this section. The results are useful in selecting the optimal discretization for a new model, or in evaluating the output of an existing model. All the formulas are derived for an arbitrary Fourier component of the pumping rate time series. The well is assumed to be circular, and situated at the center of a square cell to simplify the derivations. Even if some of these assumptions may not be true in the actual application, results of the study are useful in understanding the behavior of numerical errors near wells.

The following equation governing ground water flow around a well is used for the analysis.

$$\frac{\partial H}{\partial t} = \frac{K}{r}\frac{\partial}{\partial r}(r\frac{\partial H}{\partial r}) \tag{25}$$

Consider a solution in the form $H = R(r)T(t)$ in which $T(t) = \exp(Ift)$. Using separation of variables, (25) can be reduces to

$$r^2\frac{d^2R}{dr^2} + r\frac{dR}{dr} - \frac{IfRr^2}{K} = 0 \tag{26}$$

Using a characteristic length $\lambda = \sqrt{K/f}$, radius r can be made dimensionless as $\hat{r} = r/\lambda$. Similarly, t can be made dimensionless using $\hat{t} = ft$. The general solution of (26) that is also finite at $\hat{r} \to \infty$ can be expressed as

$$H(\hat{r}, \hat{t}) = cK_o(\hat{r}) \exp(I\hat{t}) \tag{27}$$

in which, $K_o(\hat{r})$ is a modified Bessel function; c = a constant that is different under different boundary conditions. For a sinusoidal pumping rate $Q(t)$, c can be determined by assuming that $Q(t)$ is equal to the flow rate at the boundary $\hat{r} = \hat{r}_w$ of the solution in (27). $\hat{r}_w$ = dimensionless well radius. This assumption is valid for most wells in South Florida where the storage capacity of the well is negligible. Substituting this c into (27),

$$H(\hat{r}, \hat{t}) = \frac{Q(\hat{t})K_o(\hat{r})}{2\pi K \hat{r}_w K_1(\hat{r}_w)} \sin(\hat{t}) \quad \text{for} \quad \hat{r} > \hat{r}_w \tag{28}$$

In the case of extremely small wells, $\hat{r}_w K_1(r_w) \to 1$ as $\hat{r}_w \to 0$, and (28) becomes

$$H(\hat{r}, \hat{t}) = \frac{Q_0 K_o(\hat{r})}{2\pi K} \sin(\hat{t}) \tag{29}$$

This analytical expression for pumping head shows that the amplitude decays rapidly with distance as exhibited by the behavior of K_o. Table 1 shows variation of a portion of (28) as an indicator of this amplitude. The table shows for example that when $\hat{r} > 2.75$ and $\hat{r}_w$ =0.5 or less, the amplitude of the water level fluctuation will decay to less than 5% of the amplitude at the cell containing the well. When $\hat{r}_w$ =0.1, then $\hat{r} > 1.95$ for the amplitude to decay to 5%.

In the numerical model, the governing equation and its solution (28) are represented using $\Delta x \times \Delta x$ square cells. Average values of these cells are used to represent the values in the solution. To compute the numerical error in the cell containing the well, water level at an equivalent radius to the $\Delta x \times \Delta x$ square cell is obtained assuming a solution of the form (27). This solution is compared to the water level obtained using the analytical solution given in (28) to obtain the error. The value of c is obtained using the following equation for water balance in this cell.

$$-Q(t) + 2\pi r_c K \left(\frac{\partial H}{\partial r}\right)_{r=r_c} \approx \int_{r=0}^{r_a} 2\pi r \frac{\partial H}{\partial t}\, dr \tag{30}$$

in which r_c = equivalent radius of the $\Delta x \times \Delta x$ square cell used to compute ground water seepage; r_a = equivalent radius of the square cell when used to compute water volume. r_a is estimated using $\Delta x^2 = \pi r_a^2$. Substituting the c value obtained in (30), and assuming a pumping rate of $Q(t) = Q_0 \sin(ft)$, (27) can be expressed as

$$H(\hat{r},\hat{t}) = \frac{Q_0 K_o(\hat{r}) \sin(\hat{t} - \hat{t}_o)}{2\pi K \hat{r}_c K_1(\hat{r}_c) \left(\sin(\hat{t}) - \frac{M_o(\hat{r}_a)\cos(\hat{t})}{\hat{r}_c K_1(\hat{r}_c)}\right)} \quad \text{for} \quad \hat{r} > \hat{r}_c \tag{31}$$

in which, $M_o(r_a)$ is defined as

$$M_o(\hat{r}_a) = \int_0^{\hat{r}_a} \hat{r} K_o(\hat{r})\, d\hat{r} \tag{32}$$

$\hat{t}_0$ = a time lag. The exact solution for the water level in the cell containing the well is obtained by using (28) at a radius $\hat{r} = \hat{r}_c$. This level is compared to the analytical estimate of the water level from the numerical model. This analytical estimate given by (31) at $\hat{r} = \hat{r}_c$ differs in amplitude and phase to the exact solution. The ratio between the numerical amplitude and the analytical amplitude C_c can be used to compute the error in amplitude using $\epsilon_w = 100(1 - C_c)$. Using (31) and (28), C_c can be expressed as

$$C_c = \frac{1}{\sqrt{\left[1 + \left(\frac{M_o(\hat{r}_a)}{\hat{r}_c K_1(\hat{r}_c)}\right)^2\right]}} \tag{33}$$

It can be shown that $\epsilon_w = 0$, $H_c(\hat{r},\hat{t}) = H(\hat{r},\hat{t})$ and $t_0 \to 0$ when $\hat{r}_c \to 0$. To compute the numerical values of ϵ_w approximately, $\hat{r}_c = a_c \Delta x \sqrt{(f/K)}$ and $\hat{r}_a = a_a \Delta x \sqrt{(f/K)}$ are used, based on a_c = 0.208 (Anderson and Woessner, 1991) and a_a = 0.56 obtained for steady state problems with square cells. Table 2 shows the values of ϵ_w obtained analytically for various values of $\Delta x \sqrt{(f/K)}$. Values in the table show that the amplitudes at the cell are generally a fraction of the analytical estimates. ϵ_w is the smallest numerical error possible with any time step, and depends only on the spatial discretization around the well. ϵ_w can be used to select Δx for a model. The model value of the water level is interpreted as the water level at a radial distance of $0.208\Delta x$ from the well. An approximate value of Δt for model runs can be based on a suitable value of ψ.

Table 3 shows ϵ_w values obtained using a MODFLOW model, and the corresponding analytical values obtained using (33). A 50 X 50 cell configuration with a sinusoidal pumping rate was used in the test to obtain ϵ_w numerically. About 200 pumping cycles were used to obtain the initial conditions for the test. About 200 more cycles were used to obtain the maximum amplitude. A small time step was selected as indicated by $\psi < 0.08$ to limit its contribution to the error. Table 3 shows that ϵ_w values obtained analytically and numerically using the MODFLOW model agree well, implying that the method can be used successfully to compute numerical errors in amplitude near ground water wells. It also shows that the values of $a_c = 0.208$ and $a_a = 0.56$ used are sufficiently accurate. Approximate errors further away from the well are large as a fraction of the amplitude, and small as a magnitude. These errors can be determined by adding the effects of propagation errors computed using (11).

Table 2 shows that when $\Delta x\sqrt{(f/K)}$ is larger than about 1.4, the error in the cell containing the well is more than 5%. When $\Delta x\sqrt{(f/K)} > 5$,the error is larger than 43%, and the cell size is comparatively larger than the radius of influence of the well. The dynamics of water level fluctuation in the well at this point are dominated by the storage of water in the cell. Numerical error in the amplitude of surrounding cells can be computed by considering that stress propagates outward from the center cell, and that the propagation error is given by (22). This error as a percentage is larger than the error of the center well or ϵ_w. Linear superposition can be used to compute the effects of multiple wells with steady and unsteady pumping rates.

NUMERICAL ERRORS UNDER STEADY STATE

Numerical errors under steady state conditions can be determined by using methods similar to those used under unsteady conditions. Since steady flow solutions are boundary dependent, a source term of the form $S(x,y) = 2E_0Kk^2\exp(Ikx)\ \exp(Iky)$ is used to create a disturbance in the solution of (1) far away from the boundaries that can be solved both analytically and numerically. The analytical solution of the problem can be shown to be of the form $H(x,y) = E_0\exp(Ikx)\exp(Iky)$. To obtain the numerical solution, consider an arbitrary Fourier component $H_{i,j} = E_n\exp(I\phi i)\ \exp(I\phi j)$. Substituting this component in the finite difference form of the

governing equations, and computing the ratio of amplitudes of numerical and analytical forms, an estimate for the percentage error in amplitude can be estimated as the difference between numerical and analytical solutions.

$$\epsilon_s = 100\left(1 - \frac{\phi^2}{4\sin^2(\frac{\phi}{2})}\right) \approx -100\left(\frac{\phi^2}{12} + \frac{\phi^4}{240} + \ldots\right) \tag{34}$$

in which, ϵ_s = steady state error as a percentage of the amplitude. The equation shows for example that ϵ_s exceeds 5% when ϕ exceeds 0.763. The corresponding values for 1% and 10% are 0.345 and 1.064 respectively. Equation (34) can be verified by making steady state runs for conditions with steady source terms having sinusoidal density variations. Model runs showed that the equation can be verified upto 4 decimal places of precision.

NUMERICAL ERRORS NEAR WELLS UNDER STEADY STATE

Numerical errors are large near wells because of the curvature in the solution. Thiem equation is used to compute the head distribution analytically when the water level in the cell containing the well is known. Thiem equation is expressed as

$$Q = 2\pi T_c \frac{H_2 - H_1}{\ln(r_2/r_1)} \tag{35}$$

in which, Q = pumping rate; subscripts 1 and 2 represent the well and the cell value respectively. r_2 = 0.208 Δx is use with square grids.

In order to represent numerical errors in dimensionless form, all the errors are normalized against the drawdown of the cell containing the well or the "center cell". The well is assumed to be positioned at the center of the square cell. The problem of determining the discretization then becomes a problem in geometry, in which the error in drawdown is expressed in terms of $r_I/\Delta x$ in which, r_I is the radial distance to a reference elevation or the radius of influence. The radius of influence can be computed using a number of empirical and semi-empirical equations outlined in the text by Bear (1972). For different values of $r_I/\Delta x$, the numerical error in the drawdowns of different cells including the cell containing the well can be obtained using numerical model runs. The drawdowns of different cells are measured with respect to a point at a radial

distance r_I, and the errors are computed assuming that the same drawdowns computed using (35) are exact. All errors are presented as percentages of the drawdown of the center cell, which is assumed to be equivalent to a well of diameter $0.208\Delta x$. A 50$\times$ 50 cell mesh was used to run the numerical model. Figure 6 shows the variation of the error obtained for cells at various distances. Three levels of discretization given by $r_I/\Delta x = 6$ and 14 along the axis and 7 along a diagonal are shown in the plots. All the plots in log scale follow an approximately linear behavior. If $r_I/\Delta x$ is less than about 7, the discretization is very coarse, and the error estimates may not be very reliable. The percentage error in Figure 6 can be expressed using the following approximate formula.

$$\epsilon = 2.07 \exp\left(-0.726 \frac{r}{\Delta x}\right), \quad \Delta x \leq r < r_I \tag{36}$$

in which, ϵ = error as a percentage of the drawdown in the center cell. The same equation can be written to express the absolute error as

$$H_\epsilon = 2.07 \left(\frac{Q}{2\pi T}\right) \log\left(\frac{r_I}{\Delta x}\right) \exp\left(-0.726 \frac{r}{\Delta x}\right) \quad \Delta x \leq r < r_I \tag{37}$$

These equations can also be used to obtain Δx for a model if the maximum error allowed at a distance r from the well is known.

NUMERICAL ERRORS IN THE SOURCE TERM

Rainfall and evapotranspiration are considered as source terms in the equation governing overland and groundwater flow. The source term is a major contributor to stress, mainly in regional models when far away boundaries have only a limited dynamic influence. Stresses introduced through the source term create water level variations that are subjected to errors during computations associated with the source term as well as other terms. A spatially and temporally varying rainfall pattern is used to study errors in the source term. The results for an arbitrary Fourier component of the solution are shown.

It can be shown that a solution in the complex form $H = H_0 \sin(Ikx + Iky - If_I t)$ satisfies the governing equation (1) if the source term describing rainfall excess (rainfall - evapotranspiration) is expressed as $S = s_c H_0 \sqrt{(f_I^2 + f_k^2)} \cos(kx + ky - f_I t - \gamma))$ in which k = the wave

number; f_I = frequency describing the rainfall pattern; $f_k = dKk^2$, $\gamma = \tan^{-1}(f_I/f_k)$. The above equation for H is used to obtain the exact solution when computing numerical errors during the following experiments.

An analytical expression for the numerical error created by the source term is obtained by isolating the source term first. A solution of the form $H_0 \sin(f_I t)$ satisfies the truncated equation when $S = s_c f_I H_0 \cos(f_I t)$ and diffusion terms are absent. Consider the following weighted implicit finite difference equation for the source term.

$$H_i^{n+1} = H_i^n + \frac{1}{s_c}(\alpha S_i^{n+1} + (1-\alpha) S_i^n) \tag{38}$$

Numerical error in (38) can be computed by comparing the analytical solution for H_i^{n+1}, which is $H_0 \sin(f_I t + f\Delta t)$, with its numerical solution obtained by substituting $S_i^n = s_c f_I H_i \cos(f_I t)$ in (38). After algebraic manipulations, the numerical error can be expressed as a percentage of the amplitude as

$$\epsilon_s = \sqrt{\psi_I^2 - 2\psi_I \sin\psi_I + 4\sin^2(\frac{\psi_I}{2})[1-\alpha(1-\alpha)\psi_I^2]} \tag{39}$$

in which, $\psi_I = f_I \Delta t$. For fully explicit and implicit methods, the expression reduces to

$$\epsilon_s = \sqrt{\psi_I^2 - 2\psi_I \sin\psi_I + 4\sin^2(\frac{\psi_I}{2})} \approx \frac{\psi_I^2}{2} - \frac{\psi_I^4}{72} + \ldots \tag{40}$$

The ψ_I values corresponding to 1%, 5% and 10% errors are 0.448, 0.673 and 0.802 respectively. With central differencing, these numbers become 1.073, 1.413 and 1.593 respectively. The numerical error as a result of both the source term and the diffusion term can be expressed as

$$\epsilon_{Ts} = \sqrt{\epsilon_T^2 + \epsilon_s^2} \tag{41}$$

in which ϵ_T is the error due to the diffusion terms alone, computed using (11).

The expression for total numerical error is verified by simulating the stress induced by two one dimensional rainfall patterns $N = N_0 \sin(kx - f_I t))$ traveling in opposite directions. Value of β

required to estimate ϵ_T is computed using $\beta = \psi_k/\phi^2$ in which, $\psi_k = f_k \Delta t$. Figure 7 shows the variation of the numerical error under such source induced flow for two sets of ϕ and β. More than 20 cycles of spatial waves simulated using more than 4000 time cycles were used in the experiment to obtain the results. The figure shows that the numerical and analytical estimates agree with each other.

ECONOMIC DESIGN OF MODEL DISCRETIZATIONS

The space and time discretizations used in a computer model should depend on the economics related to the costs and benefits of model runs. The cost is assumed to consist of; (1) the capital cost associated with a model setup, which is mainly the cost of a computer system with the required disk space and (2) the cost of running the model for a given time. The benefits consist of the sum of all the benefits and liabilities of knowing the water levels or the discharges with a certain accuracy. It is not easy to carry out a single cost benefit analysis for all private and public uses. However, the following description will help to understand the parameters associated with such an analysis.

The initial cost of setting up a model run includes the cost associated with the equipment needed for data storage. The time dependent portion of data volume is proportional to MNn_s in which M, N = the number of spatial discretizations in X and Y directions; n_s = the number of time slices at which data is stored. The volume of initial condition data is negligible. Since $MN = A/(\Delta x\ \Delta y)$ in which A = area covered by the model, and $n_s = T_s/\Delta t_s$, it can be shown that

$$C = c_s n_s MN = \frac{c_s T_s A f k^2}{2\psi_s \phi^2} \tag{42}$$

in which, C = data storage capacity needed for time dependent data in Bytes; c_s = data storage in bytes needed to store information about one cell for one time slice; T_s = period of simulation. $\psi_s = f\Delta t_s$ is based on the time interval Δt_s at which output data are saved. For a MODFLOW model using basic, river, drain and well packages for example, $c_s \approx$ 46 Bytes/cell/time step in single precision for the input data, and 230 Bytes/cell/time step for the output data. At a cost of \$0.33 per Mb of disk storage or \$6000 for a 18 GByte drive, the cost of setting up the model,

for input data alone, is 15 $\times 10^{-6}$ dollars/cell/time step. C in (42) can also be computed using a previous model run with known C_0, ψ_{s0} and ϕ_0.

$$C = C_0 \frac{\psi_{s0}\phi_0^2}{\psi_s \phi^2} \tag{43}$$

If for example, ϕ and ψ_s are halved, the equation shows that the storage required becomes 8 times. In equation (42), f and k correspond to the smallest details in the physical system studied using the model.

An expression for run time can be derived similarly. In order to make the result machine independent, the computational load is computed in flops instead of computer run time. The computer speed measured as the number of floating point operations per second (flops/s) can be obtained for many computers. The computer run time as a computational load for the finite difference method is (Lal, 1998)

$$t_r = c_r n_t MN = \frac{c_r T_s A f k^2}{\psi \phi^2} = \frac{c_r T_s A f k^2}{2\beta \phi^4} \tag{44}$$

in which, t_r = computational load in flops; n_t = number of computational steps; c_r = the number of floating point operations required per cell per time step. Run time for any machine can be computed when the computational load and the machine speed are known. For a Sun Sparc 20 and a Sun Ultra 2 used for the tests, the speeds are 4.1 and 13.8 Mflops respectively. Dongarra (1998) publishes a list of run times for many of the computers. Values c_r = 15.7 and 28.2 Kflops/cell/time step can be obtained for explicit and successive over relaxation (SOR) models. A value of c_r = 14.8 Kflops can be obtained for a MODFLOW model using basic, river, drain and well packages and evapotranspiration. For a MODFLOW model with 179 $\times$ 164 cells and 365 time steps for example, t_r = 159 $\times 10^9$ flops. Run time in hours is obtained by dividing t_r by 4.1 Mflops for a Sparc 10 to give 10.8 Hrs. t_r can also be obtained using t_{r0}, ϕ_0 and ψ_0 from a previous model run.

$$t_r = t_{r0} \frac{\psi_0 \phi_0^2}{\psi \phi^2} \tag{45}$$

Equation (44) for run time, and equation (11) for model error can be combined to obtain a relationship between them by eliminating β. Such a relationship is useful in understanding the cost-benefit consequences of running a model with different time steps or β. The error versus run time curve can be used with unit costs for run time and numerical error to obtain the optimal point on the curve at which the marginal gains of improved accuracy and increased run time are the same. Figure 8 shows such curves corresponding to different levels of spatial discretization representing different sizes of input data sets. The figure is obtained for the MODFLOW example with 179$\times$ 164 cells and 365 one day time steps requiring 0.5 GB of input data as a base run. The base run is based on a 5% accurate spatial representation ($\phi = 1.1$), and different computational loads correspond to different time steps or β. In the figure, a very coarse spatial discretization such as $\phi = 1.6$ needs only a small storage space, 0.24 Gb, when computed using (43). But such a model can only reach a limited level of accuracy even with a large run time. With a very fine spatial discretization such as $\phi = 0.2$, a large data storage space (15.1 Gb) is needed for the input data set, but a high level of accuracy can be reached only with a large run time. Unfortunately, unless large run times are used, it is not possible to achieve high levels of accuracy with these fine spatial discretizations.

Curves shown in Figure 8 only consider the variation of the run time and the computational error, and not the accuracy of representation of the solution. It is therefore necessary to maintain limits such as $\phi, \psi < 1.6$ to maintain an acceptable level of accuracy. When the ψ value exceeds the limit for example, the discretization can only represent Fourier components with periods ψ/ψ' times as long, in which $\psi' = 1.6$, resulting in the suppression of higher frequencies. The optimal β for equally accurate space and time representations is given by (16).

This relationship between the model error and the run time can be simplified when ϕ is small by eliminating higher order terms in the Taylor series. Assuming that $\epsilon_T \approx (fT)\, \phi^2 \beta$ for implicit 2-D models as explained in (15), this relationship becomes

$$\epsilon_T t_r = \frac{c_r T_s A f k^2\, (fT)}{2\phi^2} = \frac{t_{r0} \psi_0 \phi_0^2\, (fT)}{2\phi^2} \tag{46}$$

Figure 8 is obtained using (46) and assuming $fT = 1$. The relationship between t_r and ϵ_T in the figure is very similar to that obtained by Lal (1998) for various overland flow algorithms. To obtain an economically optimum time step for β, consider the following simple expression for the total cost of a model run.

$$P = C\ p_d + t_r\ p_r + \epsilon_T\ p_e \tag{47}$$

in which, p_d = the cost of input and output data storage in \$/Mb; p_r = cost of running the computer in \$/flop; p_e = the liability associated with making an additional error, measured in \$/ percentage point in the solution. The optimal run time and β for which the cost is minimum is the point of tangent of the error versus run time curve and the straight line describing the marginal cost line. At this point,

$$t_r = \sqrt{\frac{c_r T_s A f k^2 (fT)}{2\phi^2}} \sqrt{\frac{p_e}{p_r}} \tag{48}$$

$$\epsilon_T = \sqrt{\frac{c_r T_s A f k^2 (fT)}{2\phi^2}} \sqrt{\frac{p_r}{p_e}} \tag{49}$$

$$\beta = \sqrt{\frac{c_r T_s A f k^2}{2(fT)\phi^6}} \sqrt{\frac{p_r}{p_e}} \tag{50}$$

The fact that $c_r T_s A f k^2 = t_{r0}\psi_0\phi_0^2$ in (46) can be used to simplify the above equations when previous runs are available. If $p_e/p_r = 36$, $t_{r0}\psi_0\phi_0^2 = 1700$ Gflops, $\phi = 0.8$ and $fT = 1$ are assumed in Figure 8, $t_r = 219$ Gflops, $\epsilon_T = 6.1\%$ and $\beta = 5.5$ can be obtained using the equations. This is equivalent to a run time in a Sparc 20 of 219/4.1 = 15 Hrs. If the cost of a run is doubled, the equations show that the time step has to be increased by about 40%, if the loss of accuracy of the high frequency component can be tolerated.

APPLICATIONS

Two applications are presented in the paper. The first application demonstrates the use of the methods developed earlier to understand the results of an existing model application in South Florida. The second application shows the steps needed to select discretizations for a groundwater model.

Application to an existing model in South Florida

A number of hydrologic models are used in South Florida to solve problems of various space and time scales. These models are based on the same governing equations, and have many similar characteristics. The South Florida Water Management Model (SFWMM) (SFWMD, 1997, Fennema, et al., 1994) developed by the South Florida Water Management District (SFWMD) is one of the regional models used in the area. SFWMM is a physically based overland and groundwater flow model based on the diffusion flow assumption. It simulates flow over a very large part of South Florida. The grid used in the model is a 3.2 km (2 mile) square grid, and the time step used is 6 hrs. The time series data for the boundary conditions and the source term are provided to the model at 1 day time steps. In order to evaluate the validity of the diffusion flow assumption in the model first, consider somewhat extreme values of water depth $h = 1$ m, and slope $S_0 = 2$–5×10^{-5} in the Central and Southern Everglades during wet periods. These values can be used to compute the wave period of the shortest Fourier component that can be simulated using $T_p = 30\sqrt{(h/g)}/S_0 \approx 4$ days, as suggested by Ponce (1978). This equation is based on a maximum amplitude error of 5%. The equation shows that the diffusion assumption is valid unless events of shorter duration are simulated. If however the slopes are large, and the depths are low as in many other areas, the model can simulate events of shorter duration using finer discretizations.

The 3.2 km (2 mi) grid and the 6 hr time step in the SFWMM can represent various Fourier components in the solution with various accuracies. Table 4 shows maximum errors of type A of different Fourier components represented in the SFWMM. Subjected to a 5% maximum error, the SFWMM can represent Fourier components of wavelength as small as 18 km and period as small as 5.7 days. Using a typical high value of $K = 250\ m^2/s$ for overland flow in the Everglades, a wave length of 18 km is associated with a wave period of 2.5 days according to $f = 2Kk^2$. The time step required to represent a wave form of period 2.5 days with a maximum error of 5% is approximately 0.4 days. In the case of groundwater, $K = 8\ m^2/s$, and the wave period is 77 days. This value suggests that it is sufficient to provide time series data at 14 day intervals for

groundwater flow modeling. Any high frequency component in the ground water flow created by daily data is lost due to an error of type (B) unless more spatial grid points are added.

The numerical error in the final model output can be due to a combination of errors in various steady and unsteady state stress components. Consider a specific water level fluctuation of amplitude 1 m and period 6 days near a canal as an example. The amplitude at a distance of 6.4 km or two cells is computed by first obtaining k using $f = 2Kk^2$ as $1.557 \times 10^{-4} s^{-1}$ and then using (18). The amplitude at the distance is e^{-kx} = 0.37 m. Assuming that $\phi = k\Delta x$ = 0.5, β can be shown to be 0.52. Since $\beta >$ 0.25, it can be seen that an explicit model is unstable under the conditions. For an implicit model, the error is approximately $(1/2)kx\phi^2\beta$ or 6.5% of the amplitude, the absolute error is 6.5% of 0.37 m or 24 mm. The percentage error in discharge for this case is also approximately 6.5%. The error is largest when $fT = 1$, or at a distance of 7 km.

The error in the rain driven water level fluctuations is proportional to the rainfall intensity. In South Florida, this is the largest driving force for hydrology, and also the largest potential source of error in models. For a stationary rainfall pattern described by a period of 12 days and wave length 18 km for example, $\psi_I = f_I \Delta t$ = 0.524, and the error ϵ_s = 13.6% according to (39). For the stresses induced by this rainfall, ϕ = 1.1, and β = 0.52 which gives ϵ_T = 6.7% when using $fT = \pi/4$ and (12). The total numerical error due to both source term and diffusion term computations as a result of rain driven flow can now be computed using (41) to give 15.2%. If the wave length of the rainfall density pattern is 18 km of less, the rainfall data has to be collected with a spatial resolution of 3.2 km to maintain a <5% error in the input data set. When rainfall data is collected at a lower resolution, the model will contain only the corresponding lower frequency components. In South Florida, short duration small scale rains account for a large part of the total rain, and have to be represented accurately to increase the accuracy of model runs.

To demonstrate the use of the SFWMM to simulate water levels near a pumping well, use

Table 2 and select $\Delta x\sqrt{(f/K)} < 1.4$ for the error in the center cell to be less than 5%. With 3.2 km cells and $K = 8\ m^2/s$ for ground water, this corresponds to a pumping cycle of period $>$ 48 days. The example shows that heads computed near a well have large errors except in cased where the pumping rates change extremely slowly. Table 1 shows how the amplitude decay to less than 0.5% after 5 cells. The steady state error is less that 1.1% of the steady state drawdown of the center cell.

When the numerical error is needed at a given point in the model, the first step is to find the sources of the stresses. When the sources are found, principle of superposition can be used to find the errors due to each of the stresses assuming the governing equations to be linear. In many parts of the Everglades, stresses are mainly due to rain and canal level fluctuations resulting from operations.

An example problem showing important steps in discretization

The steps useful in determining the discretization for a new model are explained below using a 2-D groundwater model example. Assume that an implicit model is to be set up for groundwater flow. If the model is to simulate Fourier components as small as 200 m with a 1% accuracy, and that there is a specific need to limit the computational error of this component after propagating 50 m from the boundary to 50% of the local amplitude. Assume also that the disturbances are 1-D, and travel along an axis. Using $K = 100m^2/s$ for groundwater flow,

A. For a $< 0.1\%$ spatial discretization error in 1-D, pick $\phi \sim 0.16$ using (5).

B. Compute $k = 2\pi/\lambda = 2\pi/200 = 0.0314$ which gives $\Delta x = \phi/k = 5.1m$. Using $f = Kk^2$, compute $f = 0.0986s^{-1}$. This represents a wave period of 64 s.

C. Compute $fT = kX = 1.57$ as a measure of the extent of evolution of the disturbance. For a 10% error, use (12) for an implicit model to compute β as 2.38 which in turn gives Δt as 6.1 s when using $\beta = K\Delta t/\Delta x^2$.

D. Check if this Δt is small enough to represent 64 s period waves by computing $\psi = f\Delta t =$

0.6 which corresponds to 1.4% maximum error when using (5). If this is not assumed to be acceptable, Δt has to be reduced until the error becomes small.

E. Equation (44) shows that the computational load for modeling an area of 250000 m^2 for a period of 1 day is $c_r T_s A f k^2/(2\beta\phi^4)$ or 104.5 Gflops when using c_r = 14.8 Kflops. With a machine speed of 4.1 Mflops, the run time is 7.1 hrs. To reduce the run time slightly, relax the 50% error in C, and redo the steps. If the run time is extremely excessive, give up trying to represent 200 m long wave profiles with the high accuracy, and increase Δx. To get a 1 hr run time, assuming β to remain constant, $t_r \sim 1/\phi^4$ based on (16) and (45) can be used to obtain Δx = 26 m and Δt = 16 s. Step E can also be simplified using the error versus run time relationship shown in (46).

For most problems, the first step is to determine the spatial discretization required to represent the water surface profile created by the boundary or source term disturbance. The lower bound of the numerical error gets fixed once the spatial discretization is fixed. Time step for the problem is selected to make sure that the run time and the numerical error for the problem are at a proper balance. The equations used in step C have to be modified depending on the problem.

CONCLUSIONS

In this paper, the use of numerical errors and computer run times in deciding optimal spatial and temporal discretizations of overland and groundwater flow models is demonstrated. The numerical errors are classified into three different groups depending on the way they are introduced into the solution. Analytical expressions are derived to compute the maximum numerical errors and the run times of various 1-D and 2-D numerical models using non-dimensional space and time discretizations ϕ and β. Stresses and errors induced by canal level fluctuations, variable rainfall patterns and variable well pumping rates are computed. Both steady and unsteady state cases were studied. Numerical experiments were used to prove the validity of the expressions under the stated conditions.

The study shows the need to have sufficient spatial discretizations and matching temporal

discretizations if a given Fourier component is to be represented accurately in the model. It also shows that the run time and the data volume increases as the size of the this smallest Fourier component is reduced. Results show the existence of an optimal time step at which a marginal increase in run time is balanced by a marginal loss of accuracy, both measured in same units. Such an optimizations is useful specially in the case of implicit models with which a wide range of time steps can be used with no stability problems. The analytical expressions derived for numerical errors and run times are in dimensionless form so that they can be used to analyze existing models or develop future models. Some practically useful equations obtained during the study are summarized in Appendix A.

REFERENCES

Anderson, M. P. and Woessner, W. W. (1991). "Applied groundwater modeling, simulation of flow and convective transport", *Academic Press, Inc.*, New York.

Bear, J. (1979) "Hydraulics of Groundwater", *McGraw Hill*, New York.

Burden, R. L. and Fairs, J. D. (1985). *Numerical analysis*, Third Edition, Prindle, Weber & Schmidt, Boston, MA.

Dongarra, J. J. (1998). "Performance of various computers using standard linear equation software", *Computer Science Dept. University of Tennessee*, Report CS-89-85, Knoxville, TN.

Feng, Ke, and Molz, F. J. (1997). "A 2-D diffusion based wetland model", *Journal of Hydrology*, 196(1997), 230-250.

Fennema, R. J., Neidrauer, C. J., Johnson, R. A., MacVicor, T. K., Perkins, W. A. (1994). "A computer model to simulate natural Everglades hydrology", *Everglades, The Ecosystem and its restoration*, Eds. Davis, S. M. and Ogden, J. C., St. Lucie Press, FL, 249-289.

Fitz, H. C., DeBellevue, E. B., Costanza R., Boumans, R., Maxwell, T., Wainger, L. and Sklar, F. H. 91996). "Development of a general ecosystem model for a range of scales and ecosystems", *Ecological Modeling*, vol 88, 263-295.

Haitjema, H. M. (1995). "Analytic element modeling of groundwater flow", *Academic Press*, New York.

Hirsch, C. (1989). *Numerical computation of external and internal flows*, John Wiley & Sons, Inc., New York, N.Y.

Hromadka II, T. V., and Lai, Chintu, (1985). "Solving the two-dimensional diffusion flow model", *Proc. of the Specialty Conf. sponsored by the Hydraulics Div. of ASCE*, Lake Buena Vista, FL, Aug 12-17.

Lal, A. M. W. (1995). "Calibration of riverbed roughness", *J. Hydr. Div.*, ASCE, 121(9), 664-670.

Lal, A. M. W. (1998a). "Performance comparison of overland flow algorithms", *J. Hydr. Div.*,

ASCE, 124(4), 342-349.

Lal, A. M. W. (1998b). "A weighted implicit finite volume model for overland flow", *J. Hydr. Div.*, ASCE, 124(9).

McDonald, M. and Harbough, A. (1988). "A modular three dimensional finite difference groundwater flow model", Report 06-A1, *US Geological Survey, Reston, VA*.

Richtmyer, R. D. and Morton, K. W. (1967). *Difference methods for initial value problems*, Second Edition, J. Wiley and Sons, New York, NY.

SFWMD, (1997). "Documentation of the South Florida Water Management Model", Draft Document, South Florida Water Management District, West Palm Beach, FL.

Strack, O. D. L. (1989) Groundwater Mechanics, *Prentice Hall*, NJ. Townley, L. R. (1995). "The response of aquifers to periodic forcing", *Advances in Water Resources*, vol 18, 125-146.

APPENDIX A

A Summary of practically useful equations

Table A.1: Practically useful formulas in approximate form. In the equations, f = frequency of the disturbance in water level; $K = T_c/s_c$ = transmissivity/storage coefficient for ground water flow.

Equation	Reference
$X\sqrt{f/K} = 4.3$	X is the distance at which a 1-D disturbance of frequency f would decay to 5% of the amplitude.
$\Delta x = 1.1\sqrt{K/f}$	Δx gives the spatial discretization needed to represent a water surface profile with 5% accuracy. The profile is created by a disturbance of frequency f.
$\Delta x = 0.5\sqrt{K\epsilon_d/f}$	Δx needed to represent the same spatial discretization with a ϵ_d % accuracy.
$K\Delta t/\Delta x^2 = 0.14$	Δt gives the time step needed if the numerical error is limited to 5% of the disturbing amplitude.

$\Delta x\sqrt{f/K} < 1.4$	Δx gives the size of a square cell needed to solve the amplitude of a well fluctuation with a maximum error of 5% of the local amplitude.
$\Delta x\sqrt{(f/K)} = 5$	gives the largest Δx that can be used to model a pumping well (error $>$ 40%).
$r\sqrt{f/K} = 2.75$	r is the radius at which amplitude of a well with $\hat{r}_w = 0.5$ reduces to 5% of the amplitude of the well.
$\epsilon = 2.07\exp(-0.726r/\Delta x)$	ϵ gives the numerical error of a steady state well as a percentage of the drawdown.
$C = c_s T_s A f k^2/(2\psi_s\phi^2)$	data storage capacity needed to run a model for an area A for a time period of T_s focusing on frequency f and wave number k.
$t_r = c_r T_s A f k^2/(2\beta\phi^4)$	run time or computational load for a mode run.

Table 1: Amplitudes of water level fluctuations given by (28) for various well sizes $\hat{r}_w = a_c \Delta x \sqrt{(f/K)}$ representing various discretizations. Values of $K_0(\hat{r})/(\hat{r}_w K_1(\hat{r}_w))$ are shown in the table.

	$\hat{r}$							
$\hat{r}_w$	0.01	0.05	0.1	0.5	1.0	2.0	5.0	10.0
0.01	4.722	3.115	2.428	0.925	0.421	0.114	0.004	1.8×10^{-5}
0.05		3.128	2.438	0.929	0.422	0.114	0.004	1.8×10^{-5}
0.1			2.463	0.938	0.427	0.115	0.004	1.8×10^{-5}
0.5				1.116	0.508	0.137	0.004	2.0×10^{-5}
2.0						0.407	0.013	6.0×10^{-5}

Table 2: Variation of ϵ_w, the error in amplitude as a percentage of the exact amplitude, and C_c, with dimensionless Δx for a square mesh.

$\Delta x\sqrt{(f/K)}$	0.01	0.02	0.05	0.1	0.2	0.5
C_c	1.0000	0.9999	0.9997	0.9990	0.9967	0.9844
ϵ_w (%)	4.200 10^{-7}	5.130 10^{-5}	1.351 10^{-4}	1.508 10^{-3}	1.564 10^{-2}	0.289

$\Delta x\sqrt{(f/K)}$	1	2	5	10	20
C_c	0.9523	0.8886	0.5640	0.2563	4.3228 10^{-2}
ϵ_w (%)	2.13	11.1	43.6	74.4	95.7

Table 3: Comparison of values of ϵ_w obtained analytically and using the MODFLOW model.

$\Delta x\sqrt{(f/K)}$	0.54	1.53	6.10	8.63
ϵ_w (analytical)	4%	6%	70%	79%
ϵ_w (MODFLOW)	4%	5%	71%	82%
ϕ	0.27	0.76	3.05	4.31

Table 4: Characteristics of various 1-D and 2-D harmonics that can be represented by using a 3.2 km (2 mile) grid cell and a 1 day time step.

Wave length (1-D) (km)	41	18	13	6	4	19
Wave length (2-D) (km)	57	25	18	8	5	27
Max. error (%)	1%	5%	10%	50%	100%	4.5%

Wave period (Days)	12.8	5.7	4.1	1.8	1.3	6
Max. error (%)	1%	5%	10%	50%	100%	4.5%

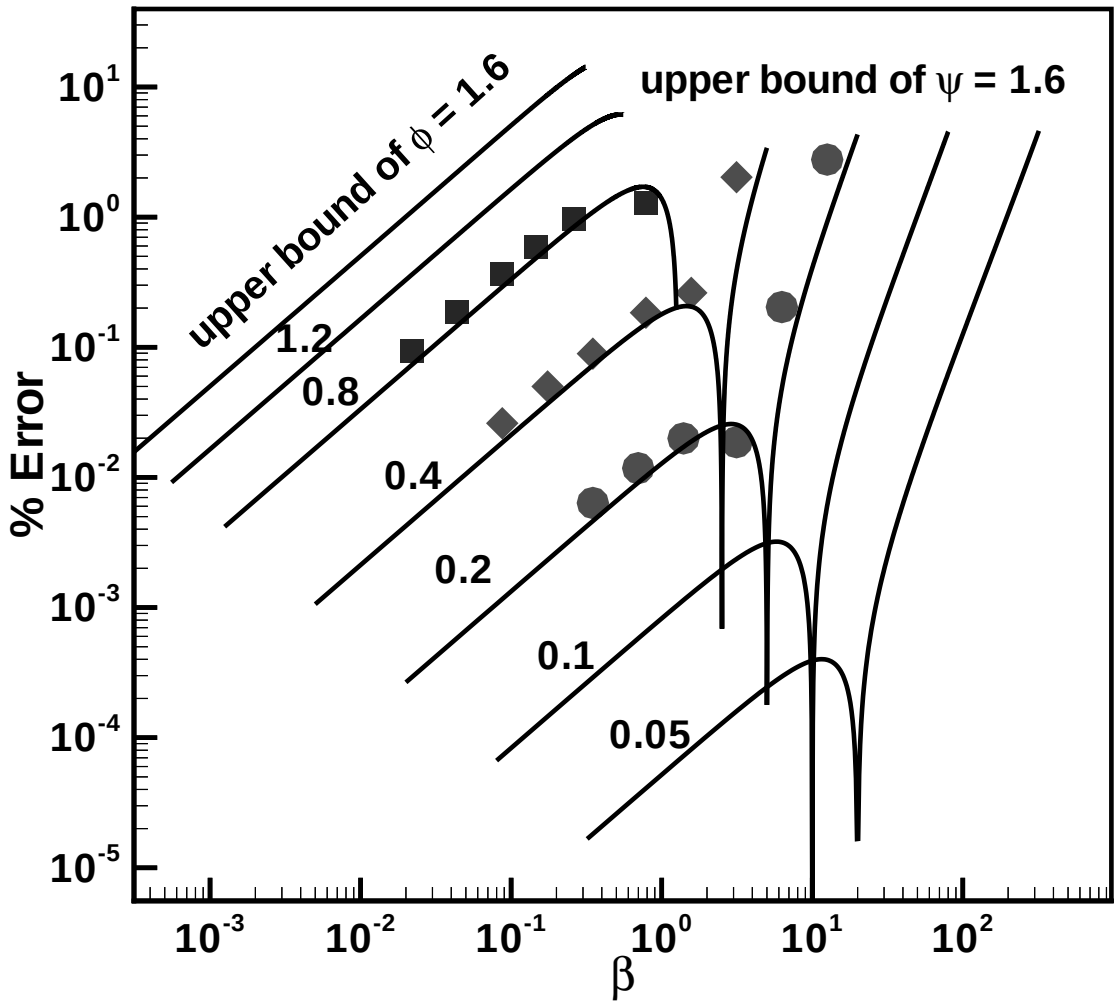

Figure 1: Variation of numerical error with spatial and temporal resolutions for the ADI method

/usr2/wlail/resolution/exper_xp.lay

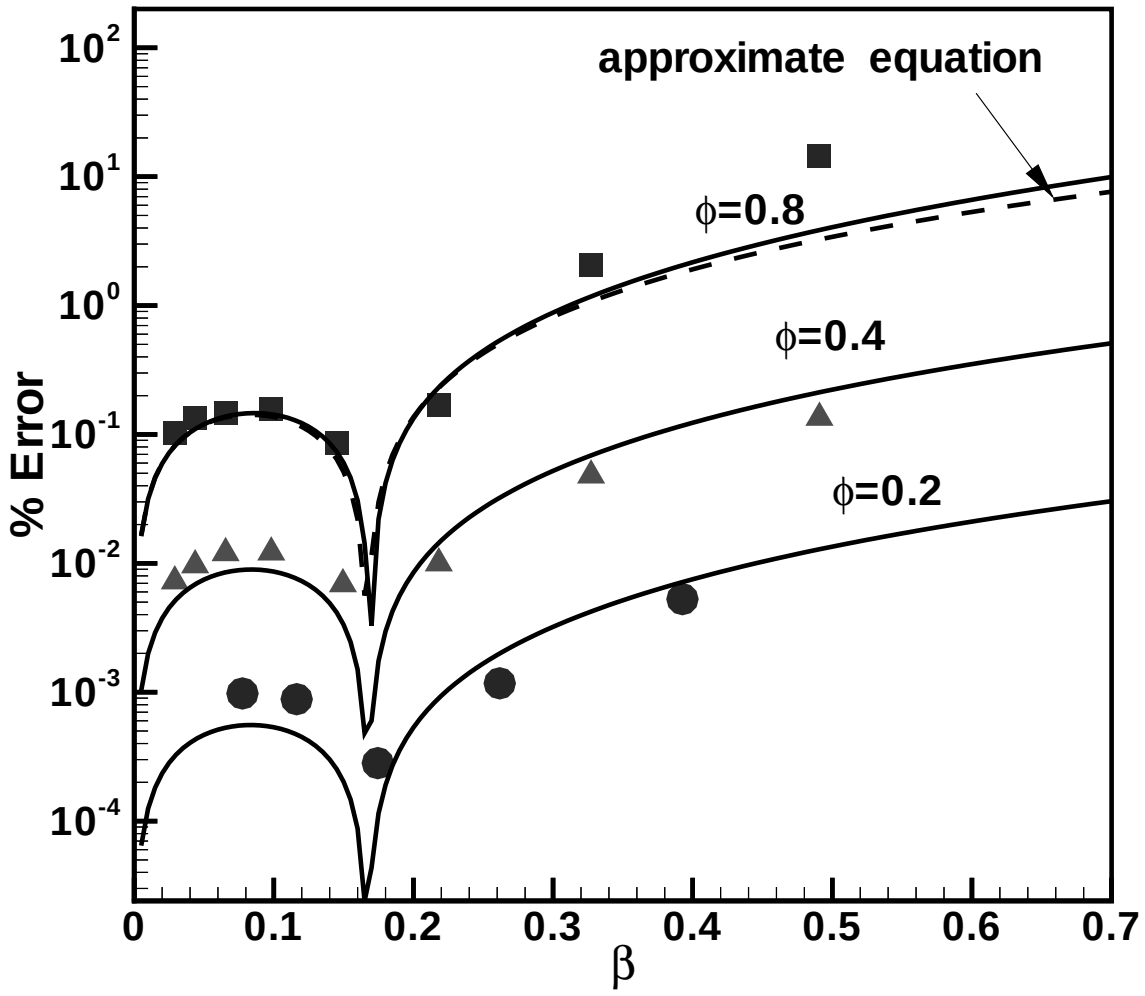

Figure 2: Variation of numerical error with spatial and temporal resolution for the explicit method.

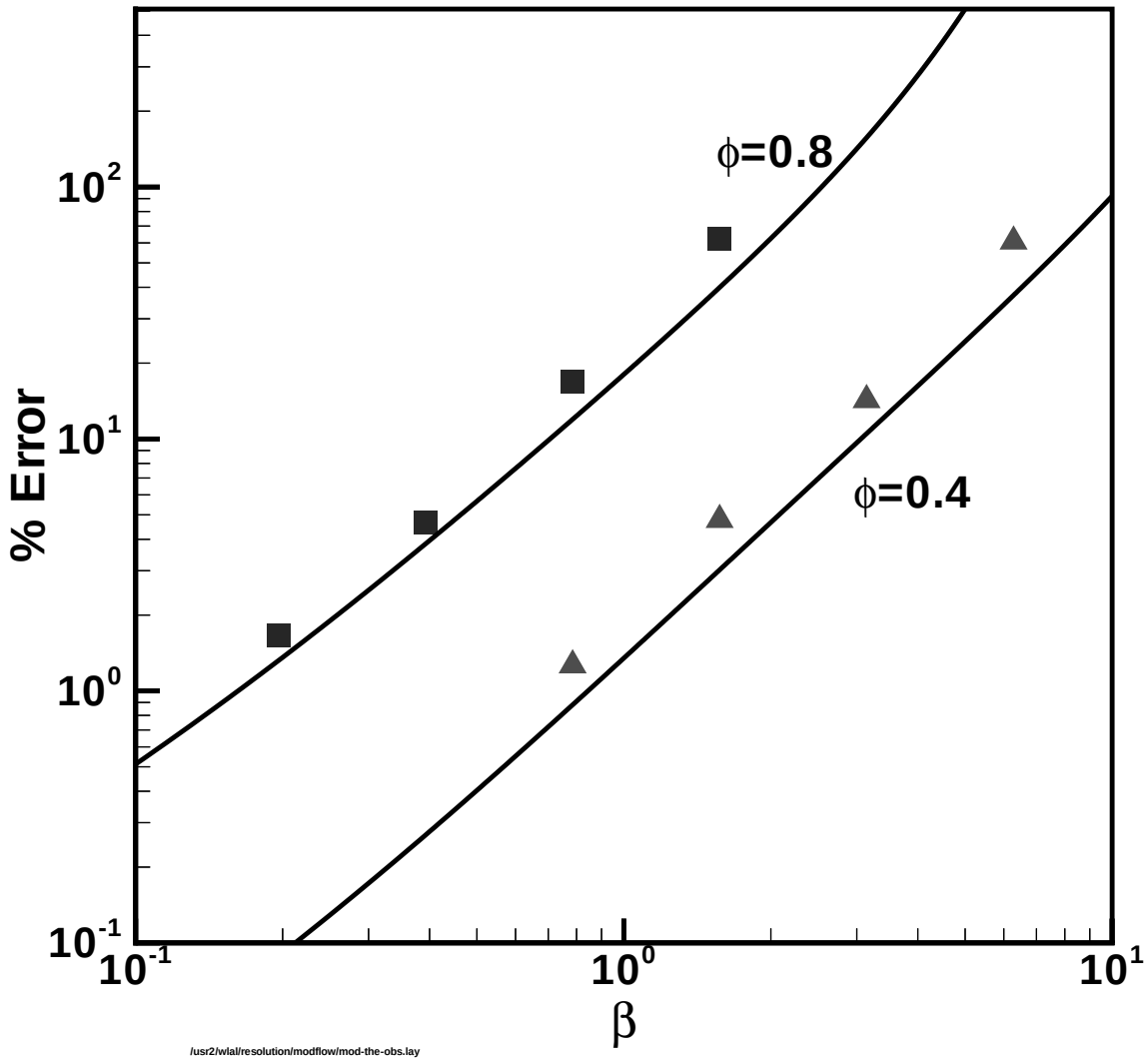

Figure 3: Variation of numerical error with spatial and temporal resolutions for the MODFLOW model.

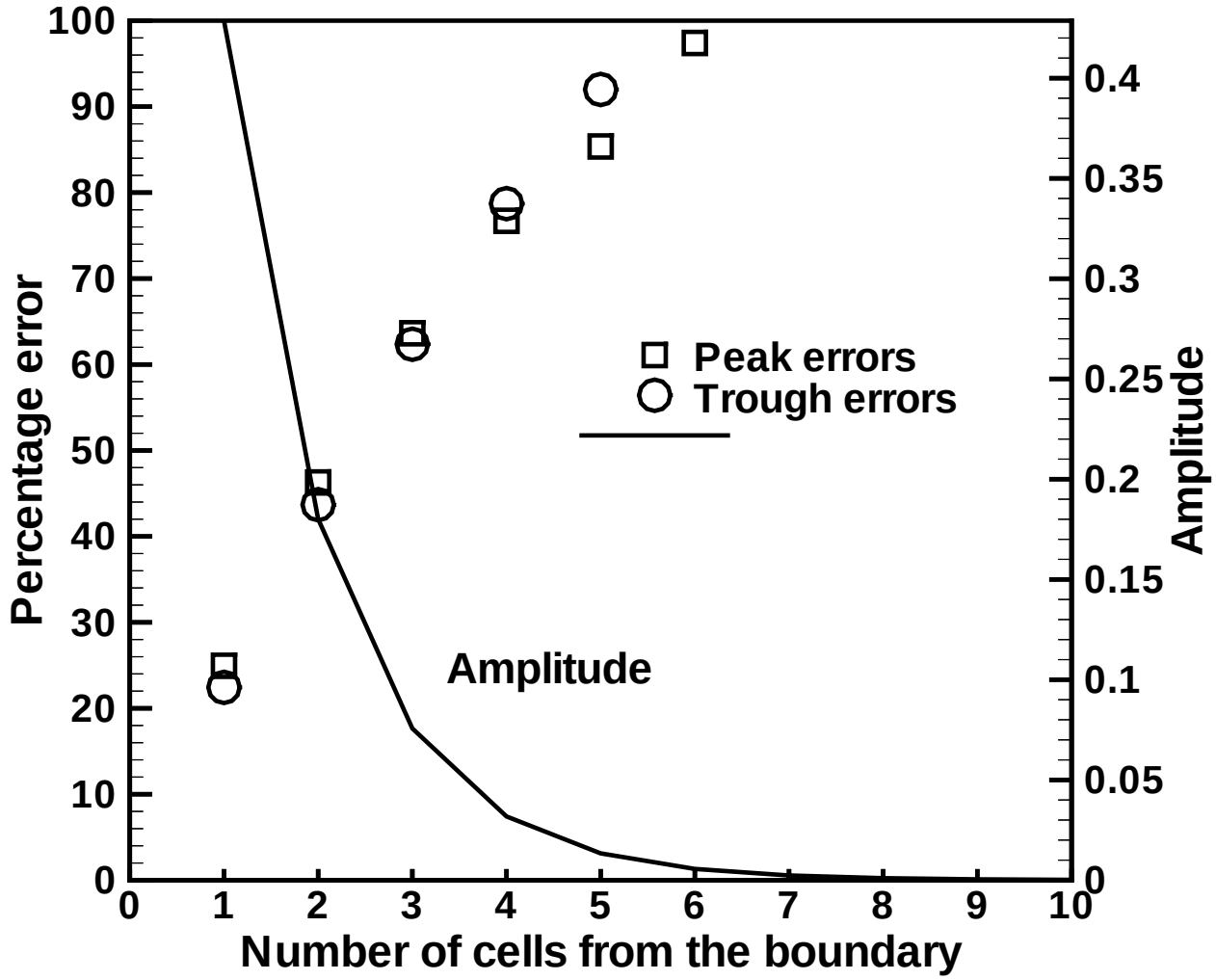

Figure 4: Variation of numerical error with distance for the MODFLOW model.

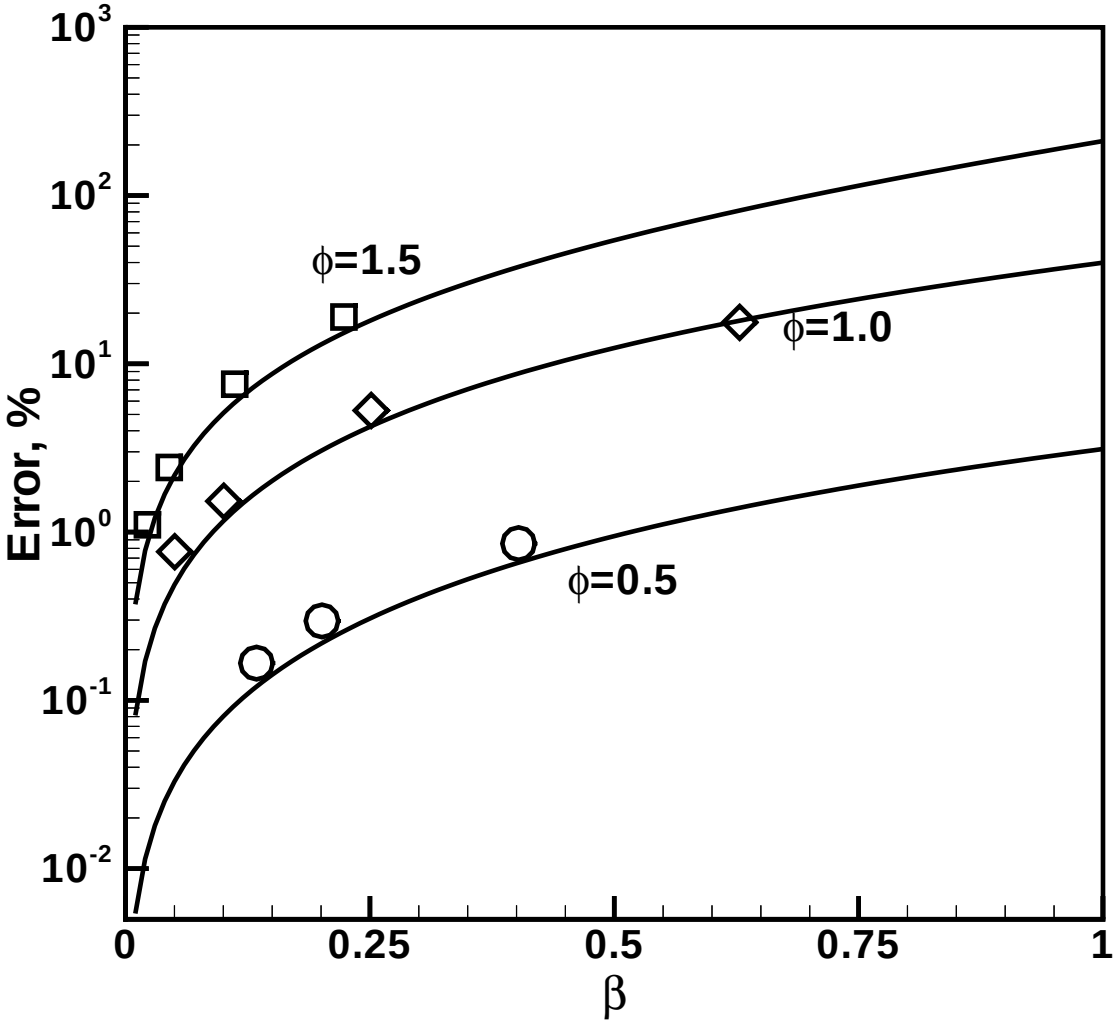

Figure 5: Variation of error in discharge for a fully implicit model.

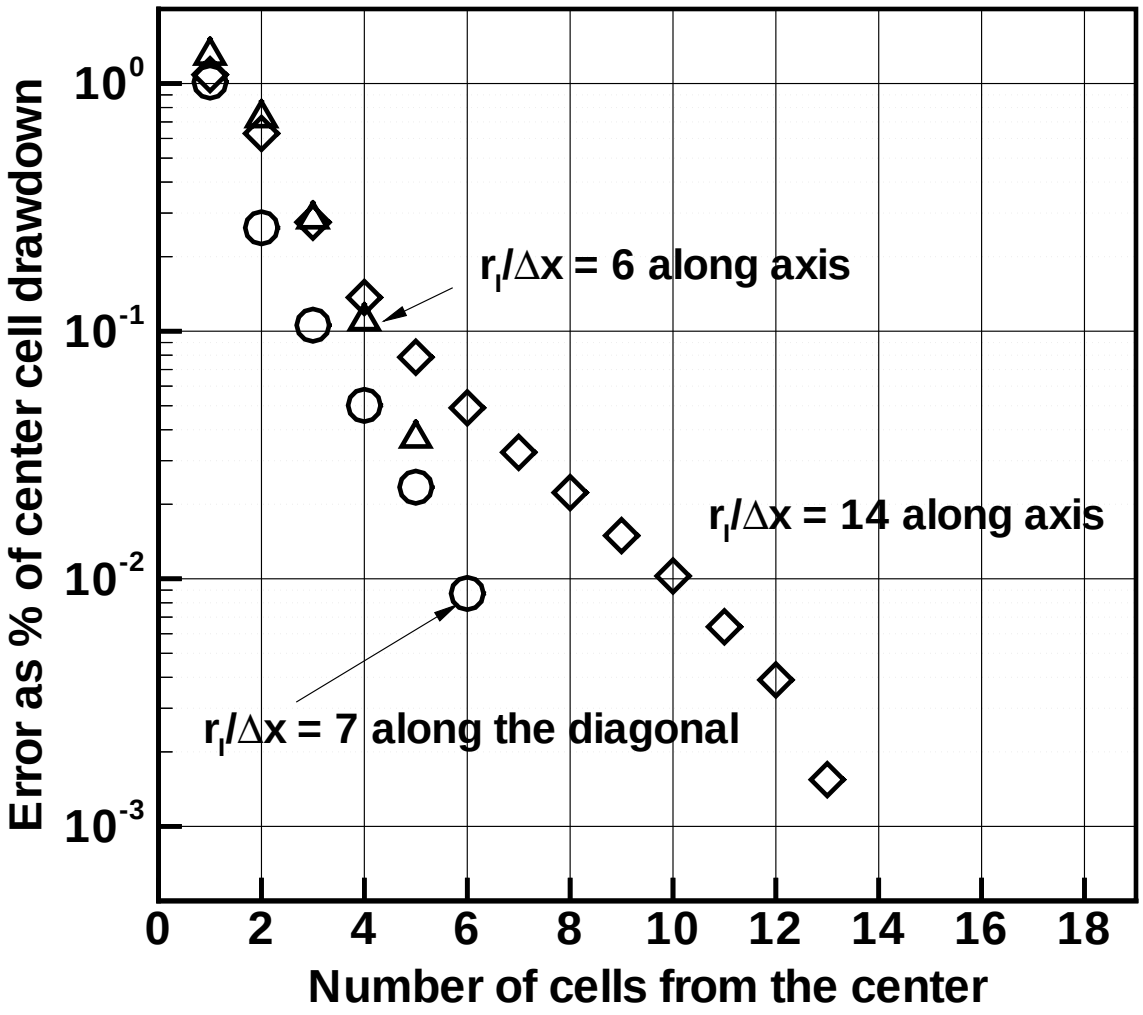

Figure 6: Variation of numerical error with radial distance.

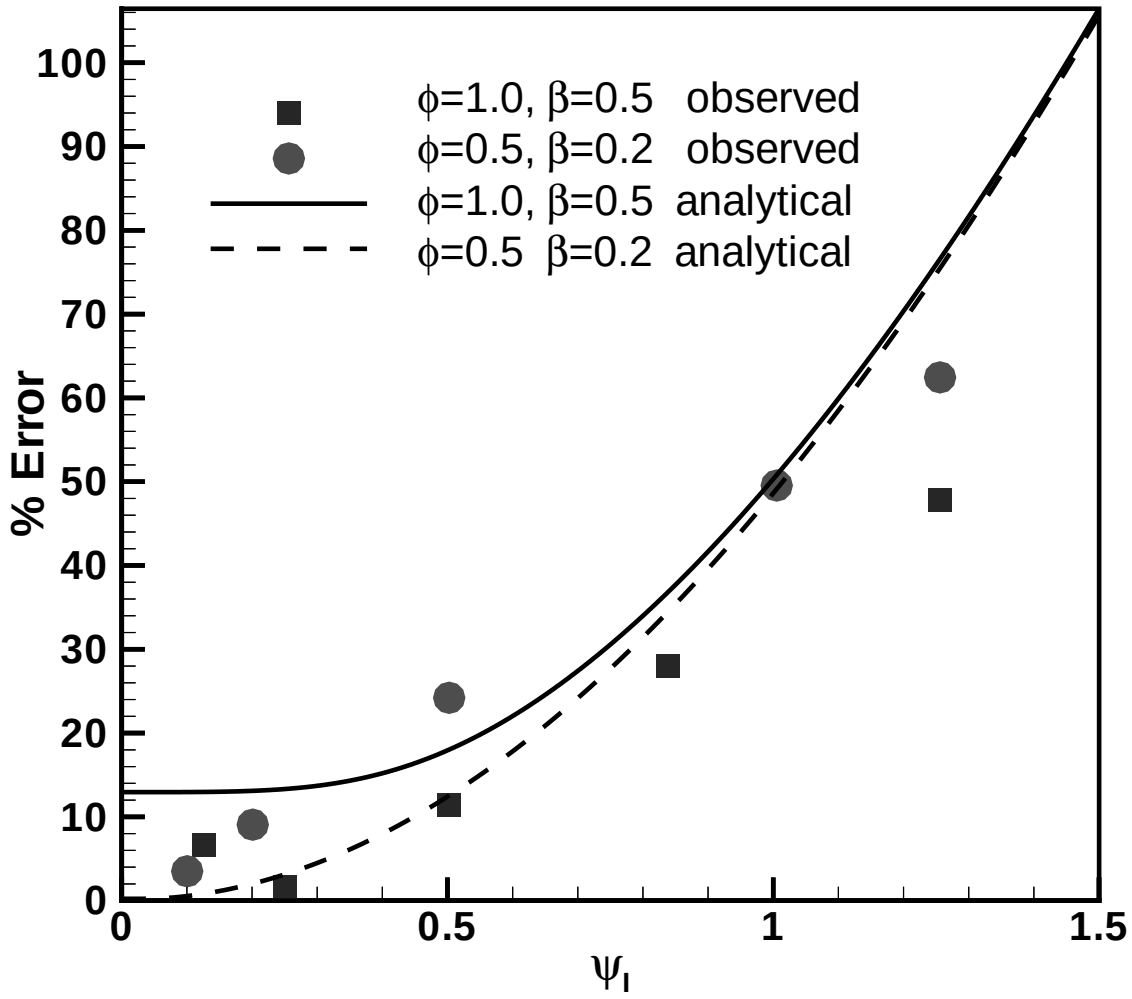

Figure 7: Variation of error with frequency of rainfall.

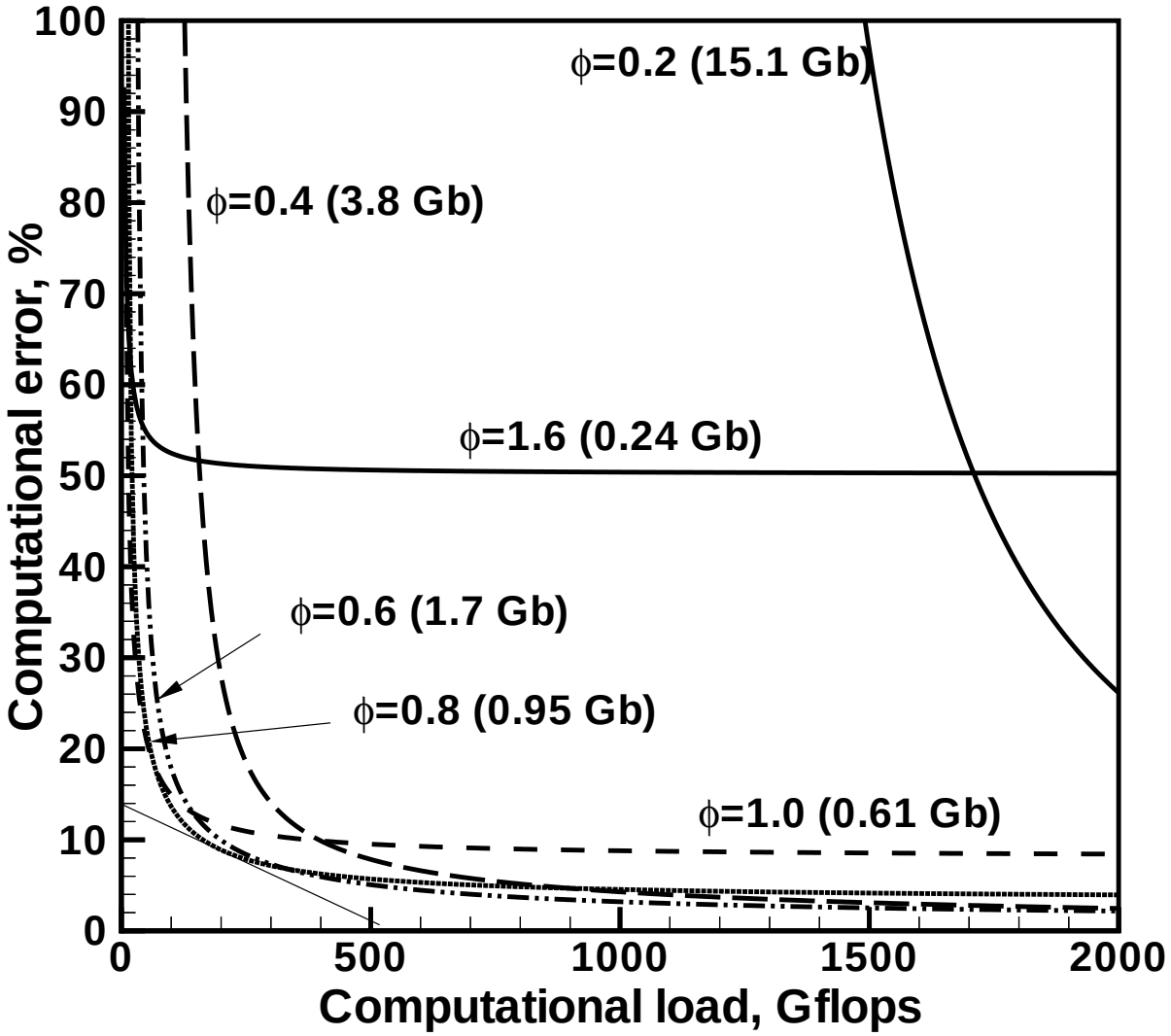

Figure 8: Variation of computational error with computational load for a 179 $\times$ 164 cell MODFLOW model running at daily time steps of 1 yr.

DEFINITION OF VARIABLES

Variable	Definition
A	area simulated by the model (m^2).
C	data storage capacity needed to run a transient model.
c_r	number of floating point operations per cell per time step (flops).
c_s	data storage needed to store information abou one cell for one time slice (Bytes).
f_I	frequency of the rainfall pattern.
g	gravitational acceleration.
h	water depth, (m).
H	water levels or water head (m).
H_ϵ	error in the steady state solution near a well.
K	T_c/s_c for groundwater flow, $h^{\frac{5}{3}}/(n_b\sqrt{S_n})$ for overland flow, m^2/s.
K_0, K_1	modified Bessel functions of type 0 and 1.
r	radial distance from the center of a well.
$\hat{r}$	$= r\sqrt{(f/K)}$ dimensionless r.
$\hat{r}_w$	well radius in dimensional form.
S	source term representing rainfall and evapotranspiration.
s_c	storage coefficient
T	time during which a given harmonic evolves, (s).
T_c	transmissivity, m^2/s
T_s	period of simulation of the model.
t_r	computational load or computer run time for a model.
x, y, z	distances along x, y, z coordinate axes, (m).
X	distances at which a disturbance is measured, (m)
α	time weighting factor in the weighted implicit scheme.

Variable	Definition
β	$K\Delta t/\Delta x^2$, dimensionless time step.
ΔA	area of a cell, (m^2).
Δt_s	time interval at which the output data is saved (s).
Δx	size of a square cell.
ϵ_Q	numerical error in discharge as a percentage of discharge.
ϵ_s	numerical error due to the computations associated with the source term.
ϵ_T	numerical error as a fraction of the local amplitude.
ϕ	dimensionless spatial discretization defined as $k\Delta x$.
ψ	a dimensionless time discretization defined as $f\Delta t$.
ψ_I	a dimensionless time discretization defined as $f_I\Delta t$.
ψ_s	a dimensionless time discretization defined as $f\Delta t_s$.